End-to-End Success Landscape in Long-Horizon Tasks

Observation point

Longer horizon
-> task-level success drop
becomes more visible

X-axis: Episode-level success
Y-axis: Effective horizon (number of episodes)
Colorbar: Estimated task-level success